Metrics and Processes for Evaluating Scholarship in Higher Education: Experts' Critical Reflections on Key Questions

Michael Saren, Montserrat Díaz-Méndez, and Evert Gummesson

Abstract In this final chapter the editors report the results of the international survey of senior academics views about the evaluation of scholarship in HE which they conducted for this book. We recruited a broad international group of senior university staff who have been involved with the evaluation of academic work in a variety of institutions, disciplines and roles over many years. We contacted over 20 experts, most of whom agreed to help and share their experience and expertise. The experts were asked for their written responses and critical reflections on five key questions about the topic. Their replies suggest that there is no one agreed best measure of academic, teaching or research quality; different indicators highlight different aspects of performance. Some respondents were concerned that there is a tendency for an "uncritical realism" towards quantifiable metrics for evaluations of teaching and research and academic work generally and an assumption that they are objective and based on merit. Regarding the vexed topic of teaching quality assessment by students, one major concern was the use and role of 'customer satisfaction' type of student surveys. Several positive features and advantages of employing appropriate metrics for academic evaluation were identified. The methodologies are generally transparent so academics being assessed cannot say they do not know the 'rules of the game'. Employing international standards to evaluate scholarship exposes academics to a competitive environment fostering their professional development and international research. Viewing the student role primarily as that of "customers" was overwhelmingly criticized, however. Although students are 'co-creators' of learning, most

M. Saren (✉)
University of Birmingham, Birmingham, UK
e-mail: majs1@leicester.ac.uk

School of Business, University of Leicester, Leicester, UK

M. Díaz-Méndez
University of Extremadura, Badajoz, Spain
e-mail: mdmendez@unex.es

E. Gummesson
University of Stockholm, Stockholm, Sweden
e-mail: eg@sbs.su.se

HE institutions now tend to address them as 'customers' in their communications, branding and marketing.

Keywords Survey questions · Evaluation · Metrics · Experts reflections

Questions for Experts

1. Please outline the main issues you think are faced by Higher Education (HE) today in evaluating scholars' research performance, quality & impact. Outline those methods, metrics and criteria you consider best practice for research evaluation.
2. How would you summarize the major strengths and any weaknesses of the metrics and processes for measuring scholarly work (teaching/research/other) that are used in the countries & cultures with which you have most experience?
3. In general, terms what should be the students' role in HE—partners, participants, customers, outputs, financiers? In addition, more specifically please expand on the part you think students should play in the assessment of teaching quality and the appropriate contribution of students' evaluation surveys.
4. The so-called marketization of HE reinforces a producer-to-consumer view of value provision. In what respects, if at all, would taking the perspective of value co-creation in the wider HE context provide a viable alternative approach for conceptualizing and evaluating scholarly work?
5. Are there any other new or alternative theoretical and methodological developments you would like see applied to the criteria or measures for evaluation of scholarly work– teaching, research, management, other?

Survey Results

Q1. The Evaluation of Research. What are the main issues faced by HE today in evaluating scholars' research performance, quality & impact. What methods, metrics and criteria do you consider best practice in research evaluation?

Most of the experts surveyed reported that research by scholars is usually assessed using the impact of their publications based on journals´ rankings, number of citations, H factor, etc. The main mechanisms used by universities for evaluating their research performance, quality and impact were those international ranking institutions that are generally recognised as reference for such attributes, such as Scimago, Scopus, QS, etc.

There were several issues regarding this methodology raised by the experts. One was whether these mechanisms and rankings are sufficient or fair measures for the task of evaluating individual' For example one view was that "evaluating scholars should never only look at figures and indexes"; these should only be used as preliminary or complementary indicators in the evaluation of how the scholar has contributed theoretically, methodologically, managerially etc. Another concern is that the quantity of publications used for the evaluation by HE institutions or State authorities in some countries can overshadow their quality and the scientific, policy or managerial

contribution. Another challenge for individual evaluation is the preponderance of multi-authored research outputs and publications. The more scholars research and publish in teams, the more difficult it is for institutional or external authorities to disaggregate the extent and nature of individual author's contributions to specific pieces of work. Some journals and universities tackle this by requiring statements of contributions per publication, signed by all authors. However, some respondents remain skeptical about the "self-rating" component of this practice as a solution for achieving accurate evaluations of research by the individual scholars. One respondent remarked, "It's like students marking their own work",

Some were concerned was that there is a tendency for an "uncritical realism" towards quantifiable metrics and an assumption that they are objective and based on merit. One respondent illustrated an example of the problem with this form of understanding here citing empirical evidence showing how gender influences citations. The assumptions underpinning this way of thinking about metric evaluation of research are not recognized in most disciplines and policy debates, yet despite these forms of research assessment are being used to produce a granularity of outcomes that they cannot achieve with sufficient accuracy.

Another set of intra-institutional problems were raised between various organizational levels carrying out research evaluations. In the evaluation of individuals' performance, the assessments by subject experts in department level having the best knowledge of the discipline can be overruled or moderated at School, College and University level by other academics having the least knowledge of the discipline. In such situations University, board tends to have to rely on the quantitative metrics supplemented by on the external reviewers' reports in a formal and thus "apparently more objective" manner. One respondent was extremely doubtful about this process because such evaluation decisions "always involve an aspect of 'politics. It utterly bizarre to expect that a completely neutral way of evaluating [scholars' research] is possible". A frequent concern expressed by respondents reporting these issues arising from internal evaluation processes is that its effect is to reduce the potential development of innovative or critical research that challenges mainstream trends in the discipline.

Despite their recognition of the problems with specific research evaluation methodologies and metrics, to a greater or lesser extent the experts surveyed generally recognized the need for some sort of assessment system. Two reasons were cited by one of our experts. First, for HE' accountability to the state as funder (whether directly and indirectly) to make sure the taxpayer is getting something in return for investment and to ensure that taxes and increasingly individuals' funds are spent efficiently and appropriately. Second, in order to provide a measure and mechanisms to support university HR process to manage the staff resource. As in any workplace, some form of performance measurement "however crude and judgmental" allows discussions and processes to take place. "Academia is not immune to expecting colleagues to do the job for which they are paid".

One respondent though that "any metrics and procedures can be deployed well or badly." The quant metrics tell us something about the research productivity and

impact…" however, "these should not be employed at the face value. The interpretation of any metrics + external reviewers' reports require deep knowledge of the field and its various paths and trends. This suggests that the persons doing the evaluation should be within the field themselves." Notwithstanding these problems with metric systems of evaluation, the general view of experts was that, although metrics cannot do it all, they are useful indicators when used in conjunction with qualitative evaluations of scholars' research. One common conclusion was that although there is a continual pressure to standardize the evaluation of performance, quality and impact through quantitative tools and ranking systems, these "heuristics are just that" and they need to be balanced with sound peer judgment. One expert concluded that both quantitative or qualitative measures will involve some degree of "informed subjectivity", therefore the key is to identify how HE managers and their regulators can make the research evaluation process "realistically nuanced to take into account subject or even sub-disciplinary norms, career stage and other equality related issues. So maybe best practice is in nuancing the tools which we use or are forced to use, rather than trying to find the 'right tools'—if they actually exist."

Another response was more sanguine about the prospects for developing better more accurate, fairer evaluation processes that assess academic performance that contributes to society while also reflecting research rigour and relevance. They cite example of movements, which attempt to study and monitor these issues such as Responsible Research for Business and Management, a virtual, global organization combining leading scholars, major accreditation bodies, and leading schools worldwide.[1] Also the San Francisco Declaration on Research Assessment (DORA), is a worldwide initiative covering all scholarly disciplines and key stakeholders, including funders, publishers, professional societies, institutions, and researchers. It explicitly recognizes the need to improve how researchers and scholarly research outputs are evaluated and provides a framework to explore the synergies between rigour and relevance of academic research and provides valuable guidance in evaluating research performance by detailing how to assess research on its merits. DORA also recognizes the need for a balance between quantitative indicator-driven and qualitative peer-review-driven assessment methods that are appropriate for each discipline.

Q2. Strengths and weaknesses of metrics and processes for measuring scholarly work (teaching/research/other) used in the countries & cultures you have experienced.

This question covers a wider range of scholarly work and as one respondent noted it "opens up another set of questions about measures". The results of the survey indicated that there is no one agreed best measure or indicator of academic teaching or research success, which can each be defined and measured in various ways. Different indicators highlight different aspects of performance. Therefore, several responses suggested that a portfolio of measures is needed to assess academic rigour, academic relevance, and practical relevance at different stages of an academic career.

---

[1] https://www.rrbm.network/.

Several positive features and advantages of employing appropriate measures were noted by one respondent. "We may not like them, we may argue that they or some elements of them are unfair, they are not consistently interpreted etc. but in general terms they are transparent and the measures are visible. "As academics who are being assessed" We cannot say we do not know the 'rules of the game'. We know that we are expected to publish in leading journals, we are expected to raise research funds, supervise doctoral students and student evaluations (de facto satisfaction surveys) should hit are certain threshold" Another note the benefit of using international standards and metrics to evaluate scholarship is that this exposes academics to a very competitive environment fostering their professional development and giving the opportunity to become part of wider research networks across regions and countries.

For a different country in the survey, however the expert reported that they had a national system of journal quality where each academic field categorized the top 20% of global output. Although not perfect, they consider it preferable to international standard metrics because the classification was based on a dialogue among experts in the academic fields being assessed. In addition, this methodology provided a legitimate and nationally tailored alternative to the dominant international ranking systems, which did not fit well with that country's institutional context or academic traditions.

There were other problems reported with paying too much attention to international indicators. First, his could reduce the incentives to focus on other less tangible attributes, resulting from participating in such competitive environment such as building strong networks, mentoring and guiding young researchers and faculty and engagement with practitioners. These type of research "by-products" are usually assumed to be included as part of those "hard" metrics, which may not necessarily always be the case. Second, there is a need to customize the assessment of scholarly work to ensure fairness across different disciplines (e.g., Engineering vs. Social Sciences vs. Medicine). Third, similarly, standards and metrics are not always modified according to scholars' career stage or development requirements.

Other responses were more critical about the use of metrics because they create the image that all applicants can be assessed objectively with same metrics and that metrics were objective and reliable. One said "This not the case." They can be manipulated….When used, several metrics should be used together and their quality/emphases revealed". Another respondent thought "they don't measure what they claim to and probably can't". In some national HE systems, the expert view is excluded from teaching assessment measures and the data are used in a de-contextualized way, thus assuming (unrealistically) that there is genuine equality of opportunity in learning and employment for all students. Another issue raised was that metrics and processes are generally linked to expectations and targets—"otherwise why are we measuring things?" Also some thought that there are fundamental questions to be asked about what the measures are actually capturing; issues about how achievable some of the goals\expectations are; and the 'validity' of survey based metrics such as student evaluations.

One expert explained that among the top echelons in of a discipline, publishing in elite journals is a critical high achievement measure. These achievements suggest

quality is based on the assurance provided by a journal's rigorous review process and what reviewers and editors judge to be an original academic contribution. However, some established journals are criticized for lacking academic and practical relevance and stifling innovation by researchers working in emerging areas because innovative research often challenges conventional thinking (Armstrong, 1997). Consequently, such research can face barriers in the review process because of "similarity bias," as the research does not fit the norms and practices with which reviewers are familiar. Newer sub-disciplines and cross-disciplinary areas often have many such barriers to overcome and thus their journals can take a long time to become established.

In order to gain an accurate judgment about an article's impact, one method is to undertake a longer-term assessment that examines how the research contributes to discovering and verifying knowledge in the discipline. For this, one respondent strongly supports the use of citations as an indicator of scientific contribution because a citation profile indicates the academic use of the research over time. Because citation conventions differ between disciplines, there are measures that they recommend to normalize citation impact measures, noting that the Web of Science and Scopus have introduced new tools for reporting the citation impacts of an author's publications. However, they highlight that while these new measures provide insight into authors' citation performance and impact, *"they fail to show the nature of scientific contribution directly."*

This expert made one important and informative caveat about the analysis of citation counts. Although they do reveal the pathways of authors' ideas, thus do not fully reveal how their work contributes to discovering and verifying knowledge in the discipline and across disciplines. More depth is needed to understand how a cited article influences subsequent and collective knowledge. "Rather than treating all citations as equal, they have to distinguish between different types such as 'application citations' (when authors cite an article because they use its findings, methods, or concepts), 'affirmation or negation citations' (when authors cite an article because their results confirm or negate the findings of the cited study), 'review citation's (when authors cite an article to illustrate what prior literature has studied) and 'perfunctory mentions' (when authors cite an article without really using it). Application citations reflect more important scientific contributions because they shape a research stream."

Q3. Students' Role in HE. What should be the students' role in HE—partners, participants, customers, outputs, financiers—and what part should they play in the assessment of teaching?

Some respondents appeared to be rather surprised by the first part of this question. One wrote "Students are students, not any of the above labels!" Another said, "Student's role is to be a student, since university is a learning community, and students have as specific role in it." Their role is axiomatically to be students in the learning process and the university life. Even if they pay for their education directly and have a part-role as "financiers", they should be primarily students. Some answers emphasized that universities are particular a kind of organizations with specific tasks and roles in the society. Taking a wider perspective one suggested that students could be thought

of as citizens—i.e. active participants in co-design and providing insight into learning processes.

The term "marketization" was not often used explicitly, but the trend toward viewing the student role primarily as that of "customers" was overwhelmingly criticized. Although many of the experts thought that "ideally students are co-creators of learning", several responses noted that in their experience that to a large extent students have now "bought into the idea of being customers." One thought that that HE institutions are to a significant extent responsible for this because they now address students as though they are their 'customers' tin their communications, branding and marketing. However when HE institutions and other authorities discuss student collectively as a 'group', there is an apparent inconsistency when they tacitly or sometimes even explicitly treat graduate numbers as 'outputs'. Of the education system.

In HE systems and programs where students pay more for their education, one the respondent considered that it is now "inevitable that will be certain expectation of the level and standard of 'service' they receive". They thought that the expectations of students have changed from being passive recipients of a 'product' to participative process which they perceive as being much more co-created –in content, outcomes or delivery mechanisms. A different view was expressed where, because they regard academic education as a mutual learning/participatory process one respondent consciously avoided ascribing students to the 'customer role'. Another questioned the assumption in this question that students role was 'fixed' in time; on the contrary, they thought that the student's role changes along their studying process. At the bachelor degree level students' participant role as learner is prominent, which then changes progressively towards a partner role at masters and especially doctoral level.

Another drew an analogy between students´ "rather unique role" with that of hospital's patients and clients' of a consultancy firm. Both students and patients have the right and are encouraged to participate in expressing their expectations of the service with which they are about to engage. At the same time, students/patients are expected to trust and let academics/doctors guide their educational/medical process. Similarly, to a health service, education's benefits are revealed over a period. They are not all usually apparent or even subjectively experienced during the course of study— or even immediately once, it is completed. Like the case of consultancy, students/ clients have more or less active involvement with different roles at different stages of the educational/consultancy process. This is another reason why benefits need to be measured over time. This response continued that is exacerbated by the restrictions that have been imposed on education processes and HE institutions due to Covid. This requires a deep reflection on the roles of both students and faculty, where both together need to work much more as partners with a less hierarchical view and enough flexibility to adapt their own roles along different stages of the educational process.

Answers to the second part of the question regarding a student's role in the assessment of teaching quality reported that they have a role in giving structured feedback on teaching and their own experience, Most ageed that they should have voice in the co-design and providing insight into learning processes however this needs to be

facilitated effectively rather than only through minimalistic survey data. The general view was that the course/teaching assessment by the students could offer a valuable tool for the teacher and the faculty, if carefully administered. In most institutions, students' feedback in teaching assessment is achieved by means of completion of some form of "customer satisfaction" survey. One underlying question posed here concerns the use or role of the evaluation survey and student input. "Is its purpose to improve the product\service and students experience or to monitor and highlight performance—of staff, systems and resources?". Moreover, the logic of the evaluation system suggests that if one of the outputs of a University education is a range of employable skills then arguably employers should be part of the evaluation process—and this would horrify many academics and teachers.

Whatever the format, all agree that student assessments should be used with 'care'. One concern is that the style of teaching and personality of the teacher should not become 'dictated' by the students. Another thought that "from a democratic perspective I think it is good that students can provide some teaching evaluations but it is not necessarily an indicator of quality". Another respondent reported, "I have seen how harmful anonymous on-line 'rate my professor' systems can be for staff, they is also evidence that these types of assessments are biased against women and minorities". Many of these surveys are institution wide and generic raising some issues about the relevance of specific questions in the survey in some disciplines. *It can occur with surprising frequency that students who give an instructor lower evaluations early in their course, evaluate much more highly them at the end of the program once students have realized the long term benefits generated by those same attributes they initially rated less favourably.* There were other issues raised concerning other 'technical' arguments about the use of scales, completion rates, validity of results etc. Evaluation questionnaires, as the title suggests, implies a post hoc evaluation of an experience.

For the foregoing reasons some of the information from student surveys can be used to improve elements but many instructors find the most useful feedback is often found in open-ended comments. More problematically, it is not always the case that institutions and their academic and professional support staff fully recognize the extent to which student expectations have changed. Therefore, it may be more important for surveys to focus on understanding expectations—as opposed to recording satisfaction –but equally important to ensure that students understand those HE resources are finite and not all expectations will be nor can be met.

One respondent concluded that "these days all elements of a paid service are 'rated' by the purchaser\client these days—rate my hotel room, complaint process, delivery driver etc.—so HE is unlikely to be an exception. The key question is what do these 'ratings' measure and how the rating are used." Student evaluations should be designed and used for service improvement and collected for key parts of the 'service'—content, outcomes, delivery mechanisms, engagement. From a Service Theory perspective, it should evaluate 'learning quality'—not 'teaching quality'. This encompasses much more than just teaching but the whole education environment and ecosystem.

Q4. In what respects would taking the perspective of Value Co-creation in the HE Service Ecosystem provide a better innovative approach for conceptualizing and evaluating scholarly work.

This question perplexed some of the experts surveyed. Not many had a marketing background, and even some of those who did were not au fait with the specifics of service theory, which underpin this question. Responses ranged from "this is a broad question. Not sure if I get it right" to "Not sure I fully understand the point of the question." and "I think that is a good idea, but it is a hard question.

Nevertheless, there was general agreement that this is a good idea. One university leader reported positively "This can be done—we are focused on this in our university". Their students are explicitly co-leaders, they have authored parts of the university strategy and fully collaborated in its formation and we are working with them on theirs. We jointly set priorities and they act as reverse-mentors for the university executive. "Students deserve value for money and this is best delivered by a fully engaged approach in which we all regard each other as humans who can make a contribution to more effective educational communities." Another agreed that value in HE is co-created, and the primary actors are the professors and students. Stressing that value in use is essential to get the degree, to get a job, but also to acquire competence and knowledge to be used in employment and life.. "I have hard time thinking how the value co-creation and value in use concepts would not fit to education, and the system that provides it."

Another respondent who was familiar with the service theory underpinning of the question considered that "value co-creation is an excellent platform to innovate the conceptualization and measurement of scholarly work. Resource integration would be particularly relevant aspect for such innovation." i.e. creating a learning environment where results would be bigger, more significant, than the sum of the individual parts. This would entail moving from "a passive to active to interactive approach, where evaluating scholarly work would become more like a project-type assessment, rather than just a passive grading". Along with that, the whole HE service ecosystem would need to work on its own conceptualization and measurement. Scholarly work is only one of many other assessments along the HE service ecosystem, which would need to be aligned systemically to ensure their viability.

One set of potential problems that another answer presented was the ways in which this would affect evaluations of scholarly work, other than perhaps in a collective way. Observing that most HE workers such as research groups, "operate on the basis of a division of labor (some are good at landing grants, others engage in external impact and others publish in top journals), however at the end of the day, individuals, not groups, are hired and fired". So there is a tension between collective vs individualist elements, that is built into our system, much like for example in many sports or in music an orchestra.

Another question raised concerned who are the recipients (and evaluators) of our scholarly work. Usually, evaluations are based around some measure of the 'outputs' of scholarly activity—a degree, an employee, a scientific discovery etc.

Something that has a purpose for the ordinary person. One issue here is that 'marketization' perspective should be about value co-creation rather than producer-to-consumer provision. "But 'marketization' (as conceptualized in the question) or maybe 'commercialization' of HE would open the sector up to the full force of the market which would have consequences for some areas of scholarly activity that might not be immediately perceived as valuable to a particular group of HE users." Some differences between university disciplines and departments was considered relevant to this question. For professional or 'vocational' degrees—business, engineering, medicine, law—some thought that a 'production factory' model of education has been taking over for a long time., For these subjects as other in HE, value-cocreation is considered an more appropriate model.

Another academic took a different approach by considering the relevance of this question for academic research as opposed to teaching. Arguing for a processual interactive approach here too they cite Gummesson[2] (2004 p. 317) who drew attention to the failure of academic research to bridge theory and practice, "researchers seem to settle for theory on a low level of abstraction or generality and have difficulties seeing the broader, systemic context; the core of a phenomenon is obscured by details and fragments." Also, "too much research is stuck in the middle neither being firmly based in real-world data nor reaching a sufficient level of abstraction." A process perspective would focus on interactivity and engagement with the various stakeholders and involve considering practical relevance to stakeholders other than practicing managers.

Q5. Alternative theoretical and methodological developments for the evaluation of scholarly work.

Not all respondents had any recommendations or comments on this question. Those who did expressed a general view that adversarial, abstracted forms of measurement were not appropriate for the sort of learning students require and therefore that which academics and instructors should do because they breed defensiveness and performing to the criteria rather than focusing on learning and research. One said, "The old saying that we should assess as if learning mattered is a good one". I would like universities to be learning organisations and criteria and measures should help the community develop and improve performance."

An alternative measure and methodology from the Service Management literature was another suggestion—Internal Service Quality. This is a general approach, not exclusively designed for educational institution that can be applied in all types of organizations (manufacturing, service, public, private). For all such organisations, the objective is to combine activities of work facilitation, internal service creation and delivery and service climate improvement to have a more positive impact on customer perceptions of service. In HE institutions (like others), there are many with hierarchical, silo-structures, with bureaucratic processes and clashing objectives which create obstacles to the systemic alignment of the whole ecosystem to facilitate

---

[2] Gummesson, E. (2004). "Qualitative research in Marketing: Road-map for a wilderness of complexity and unpredictability," *European Journal of Marketing*, 39 (3), 309–327.

learning and value co-creation. The designing of dedicated internal service quality measures using this approach in HE would help a more focussed and at the same time more comprehensive, evaluation of scholarly work.

Other reponses had recommendations, not for new evaluation methodologies, but for the way in which HE institution ought to approach and organise the evaluation process. One point was that whatever alternative measures and /or evaluations emerge they should be sufficiently nuanced not to be driven or benchmarked against some 'dominant' disciplines and they should specifically reflect academics' career stage. Another recommendation was that deans and directors of professional schools should think their faculties as teams of highly educated specialists with various orientations and interests. This, they thought "should lead to fine tuning the responsibilities of faculty persons in terms of their research contributions, teaching contributions and academic/societal/managerial services". All should have a solid research education. However, it was recognized that "this is a difficult 'model' as it requires very wise deans and faculty understanding the logics and having mutual respect, tolerance, and integrity".

There were further concerns about the expectations on early career or new academics, are become increasingly unrealistic in terms of measurable activities that can be evaluated, The opportunity for new academics to learn and grow into the job, which had been possible in the past, is becoming increasingly difficult now as more and more of their activities are being evaluated in some cases ever more frequently. One said "The expectation is that all appointments at starting level are 'oven-ready'; just look at the 'essential criteria' on job adverts—I'm not sure I would have met those expectations 30 odd years ago".

Innovation in techniques for tracking and evaluating research work and output was another cause of some disquiet. Developments in bibliometric methods that analyze article content are becoming more powerful and sophisticated, which use big data network science tools to take qualitative analyses of research into account. Bibliometric methods provide an opportunity for a more nuanced understanding of scientific contribution. Such methods include text analysis that can show how articles contribute to the evolution of a research stream. These type of techniques may be reasonably accurate and perhaps even become valuable if applied appropriately, but one important feature is that they focus primarily on outputs of research, not so much on inputs. The problem here is that the contribution of scientific and academic research also involves and depends on the performance of other institutional processes and academic activities which are *not measured or even taken into account;* for example, editorial work, community engagement, journal reviewing, research collaborations, research funding, and doctoral examination and supervision, all of which contribute to advancing knowledge.

In conclusion, this respondent listed some of their own key questions regarding the future evaluation of HE research work:

– How does the nature of citations differ for conceptual, empirical, and methodological articles?
– When does an article make a seminal contribution?

- Under what conditions do seminal articles have influence beyond their immediate disciplines?
- What alternative metrics could be used to develop a more sophisticated understanding of the nature of citations for seminal and other highly cited articles?